

0570  
05702

#2

OIPE

## RAW SEQUENCE LISTING

PATENT APPLICATION: US/10/086,208

DATE: 05/07/2002

TIME: 11:26:07

Input Set : N:\Crf3\RULE60\10086208.raw

Output Set: N:\CRF3\05072002\J086208.raw

## SEQUENCE LISTING

## 3 (1) GENERAL INFORMATION:

5 (i) APPLICANT: Goldstein, Gideon

7 (ii) TITLE OF INVENTION: Methods and Compositions for Impairing  
Multiplication of HIV-1

10 (iii) NUMBER OF SEQUENCES: 124

## 12 (iv) CORRESPONDENCE ADDRESS:

13 (A) ADDRESSEE: Howson and Howson

14 (B) STREET: Spring House Corporate Cntr., P.O. Box 457

15 (C) CITY: Spring House

16 (D) STATE: PA

17 (E) COUNTRY: USA

18 (F) ZIP: 19477

## 20 (v) COMPUTER READABLE FORM:

21 (A) MEDIUM TYPE: Floppy disk

22 (B) COMPUTER: IBM PC compatible

23 (C) OPERATING SYSTEM: PC-DOS/MS-DOS

24 (D) SOFTWARE: PatentIn Release #1.0, Version #1.30

ENTERED

## 26 (vi) CURRENT APPLICATION DATA:

27 (A) APPLICATION NUMBER: US/10/086,208

28 (B) FILING DATE: 28-Feb-2002

29 (C) CLASSIFICATION:

## 31 (vii) PRIOR APPLICATION DATA:

33 (A) APPLICATION NUMBER: US/09/451,067

34 (B) FILING DATE:

35 (A) APPLICATION NUMBER: 09/113,921

36 (B) FILING DATE:

## 38 (viii) ATTORNEY/AGENT INFORMATION:

39 (A) NAME: Bak, Mary E.

40 (B) REGISTRATION NUMBER: 31,215

41 (C) REFERENCE/DOCKET NUMBER: GGP2AUSA

## 43 (ix) TELECOMMUNICATION INFORMATION:

44 (A) TELEPHONE: 215-540-9200

45 (B) TELEFAX: 215-540-5818

## 48 (2) INFORMATION FOR SEQ ID NO: 1:

## 50 (i) SEQUENCE CHARACTERISTICS:

51 (A) LENGTH: 72 amino acids

52 (B) TYPE: amino acid

53 (C) STRANDEDNESS:

54 (D) TOPOLOGY: linear

## 56 (ii) MOLECULE TYPE: protein

## 61 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

63 Met Glu Pro Val Asp Pro Arg Leu Glu Pro Trp Lys His Pro Gly Ser

RAW SEQUENCE LISTING  
PATENT APPLICATION: US/10/086,208

DATE: 05/07/2002  
TIME: 11:26:07

Input Set : N:\Crf3\RULE60\10086208.raw  
Output Set: N:\CRF3\05072002\J086208.raw

|     |                                                                 |     |    |    |
|-----|-----------------------------------------------------------------|-----|----|----|
| 64  | 1                                                               | 5   | 10 | 15 |
| 66  | Gln Pro Lys Thr Ala Cys Thr Asn Cys Tyr Cys Lys Lys Cys Cys Phe |     |    |    |
| 67  | 20                                                              | 25  | 30 |    |
| 69  | His Cys Gln Val Cys Phe Thr Thr Lys Gly Leu Gly Ile Ser Tyr Gly |     |    |    |
| 70  | 35                                                              | 40  | 45 |    |
| 72  | Arg Lys Lys Arg Arg Gln Arg Arg Arg Ala Pro Gln Asp Ser Gln Thr |     |    |    |
| 73  | 50                                                              | 55  | 60 |    |
| 75  | His Gln Val Ser Leu Ser Lys Gln                                 |     |    |    |
| 76  | 65                                                              | 70  |    |    |
| 78  | (2) INFORMATION FOR SEQ ID NO: 2:                               |     |    |    |
| 80  | (i) SEQUENCE CHARACTERISTICS:                                   |     |    |    |
| 81  | (A) LENGTH: 912 base pairs                                      |     |    |    |
| 82  | (B) TYPE: nucleic acid                                          |     |    |    |
| 83  | (C) STRANDEDNESS: double                                        |     |    |    |
| 84  | (D) TOPOLOGY: linear                                            |     |    |    |
| 86  | (ii) MOLECULE TYPE: cDNA                                        |     |    |    |
| 89  | (ix) FEATURE:                                                   |     |    |    |
| 90  | (A) NAME/KEY: CDS                                               |     |    |    |
| 91  | (B) LOCATION: join(1..876, 883..912)                            |     |    |    |
| 94  | (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:                        |     |    |    |
| 96  | GAG CTC TAC AAA TCC GGG GAT CCG GGT GAA GAT CCG CGT TTA GAG CCG | 48  |    |    |
| 97  | Glu Leu Tyr Lys Ser Gly Asp Pro Gly Glu Asp Pro Arg Leu Glu Pro |     |    |    |
| 98  | 1 5 10 15                                                       |     |    |    |
| 100 | TGG AAA CAC CCG GGT TCT GGT TCT GTC GAC CCT AAC CTT GAA CCT TGG | 96  |    |    |
| 101 | Trp Lys His Pro Gly Ser Gly Ser Val Asp Pro Asn Leu Glu Pro Trp |     |    |    |
| 102 | 20 25 30                                                        |     |    |    |
| 104 | AAG CAT CCT GGC AGC TCC GGA GTC GAT CCC AAA CTC GAG CCC TGG AAA | 144 |    |    |
| 105 | Lys His Pro Gly Ser Ser Gly Val Asp Pro Lys Leu Glu Pro Trp Lys |     |    |    |
| 106 | 35 40 45                                                        |     |    |    |
| 108 | CAC CCC GGA AGT TCG GGG GTA GAC CCA TCT CTG GAA CCA TGG AAG CAT | 192 |    |    |
| 109 | His Pro Gly Ser Ser Gly Val Asp Pro Ser Leu Glu Pro Trp Lys His |     |    |    |
| 110 | 50 55 60                                                        |     |    |    |
| 112 | CCA GGG AGT GGT AGC GTG AAT CCG TCA TTA GAG CCG TGG AAA CAC CCG | 240 |    |    |
| 113 | Pro Gly Ser Gly Ser Val Asn Pro Ser Leu Glu Pro Trp Lys His Pro |     |    |    |
| 114 | 65 70 75 80                                                     |     |    |    |
| 116 | GGT TCA TCT GGA GTT GAT CCT CGC TTG GAA CCT TGG GAG CAT CCT GGT | 288 |    |    |
| 117 | Gly Ser Ser Gly Val Asp Pro Arg Leu Glu Pro Trp Glu His Pro Gly |     |    |    |
| 118 | 85 90 95                                                        |     |    |    |
| 120 | TCG TCC GGT GTA GAC CCC CGA CTT GAG CCC TGG AAT CAC CTC GGG AGT | 336 |    |    |
| 121 | Ser Ser Gly Val Asp Pro Arg Leu Glu Pro Trp Asn His Leu Gly Ser |     |    |    |
| 122 | 100 105 110                                                     |     |    |    |
| 124 | TCA GGC GTA GAT CAT CGG CTC GAA CCA TGG AAA CAT CCA GGT TCT GGA | 384 |    |    |
| 125 | Ser Gly Val Asp His Arg Leu Glu Pro Trp Lys His Pro Gly Ser Gly |     |    |    |
| 126 | 115 120 125                                                     |     |    |    |
| 128 | GAT CTG CGC CAG CGG CGA CGT ACT CCT CAG GAT TCT GGA TCT CGA CAA | 432 |    |    |
| 129 | Asp Leu Arg Gln Arg Arg Arg Thr Pro Gln Asp Ser Gly Ser Arg Gln |     |    |    |
| 130 | 130 135 140                                                     |     |    |    |
| 132 | CGT CGG CGC CCT CCC CAA GAC TCC TCA GGA CGG CAG CGC CGA CGA CCC | 480 |    |    |
| 133 | Arg Arg Arg Pro Pro Gln Asp Ser Ser Gly Arg Gln Arg Arg Arg Pro |     |    |    |

## RAW SEQUENCE LISTING

PATENT APPLICATION: US/10/086,208

DATE: 05/07/2002

TIME: 11:26:07

Input Set : N:\Crf3\RULE60\10086208.raw  
 Output Set: N:\CRF3\05072002\J086208.raw

|     |                                                                 |     |     |     |     |
|-----|-----------------------------------------------------------------|-----|-----|-----|-----|
| 134 | 145                                                             | 150 | 155 | 160 |     |
| 136 | CCA CAG GGT TCA GGT TCA CGT CAA CGA CGC GGT CCA CCC CAA GGC TCG |     |     |     | 528 |
| 137 | Pro Gln Gly Ser Gly Ser Arg Gln Arg Arg Gly Pro Pro Gln Gly Ser |     |     |     |     |
| 138 | 165                                                             | 170 | 175 |     |     |
| 140 | GGT TCG CGC CAG CGG CGA CGT CCG CCT CAG AAC TCT AGT GGA CGA CAA |     |     |     | 576 |
| 141 | Gly Ser Arg Gln Arg Arg Pro Pro Gln Asn Ser Ser Gly Arg Gln     |     |     |     |     |
| 142 | 180                                                             | 185 | 190 |     |     |
| 144 | CGT CGG CGC TCT CCC CAA GAT TCC GGC GGG CGG CAG CGC CGT CGA TCA |     |     |     | 624 |
| 145 | Arg Arg Arg Ser Pro Gln Asp Ser Gly Gly Arg Gln Arg Arg Arg Ser |     |     |     |     |
| 146 | 195                                                             | 200 | 205 |     |     |
| 148 | CCA CAG AAC TCA GGT GGG CGT CAA CGA CGC CGG ACT CCG CAA TCT TCA |     |     |     | 672 |
| 149 | Pro Gln Asn Ser Gly Gly Arg Gln Arg Arg Arg Thr Pro Gln Ser Ser |     |     |     |     |
| 150 | 210                                                             | 215 | 220 |     |     |
| 152 | TCC GGC CGC CAG CGG CGA CGT GCC CAT CAG AAT AGC GGC AGC CGA CAA |     |     |     | 720 |
| 153 | Ser Gly Arg Gln Arg Arg Ala His Gln Asn Ser Gly Ser Arg Gln     |     |     |     |     |
| 154 | 225                                                             | 230 | 235 | 240 |     |
| 156 | CGT CGG CGC GCA CAC CAA GAC AGC AGT GGG CGG CAG CGC CGT CGA GCG |     |     |     | 768 |
| 157 | Arg Arg Arg Ala His Gln Asp Ser Ser Gly Arg Gln Arg Arg Arg Ala |     |     |     |     |
| 158 | 245                                                             | 250 | 255 |     |     |
| 160 | CCT GAA GAT AGT GGT TCT CGT CAA CGA CGC CGG GCT CCC CCT GAC AGC |     |     |     | 816 |
| 161 | Pro Glu Asp Ser Gly Ser Arg Gln Arg Arg Arg Ala Pro Pro Asp Ser |     |     |     |     |
| 162 | 260                                                             | 265 | 270 |     |     |
| 164 | TCC GGA CGC CAG CGG CAA CGT GCA CCA GAT AGT TCC TCA GGT CAT CAC |     |     |     | 864 |
| 165 | Ser Gly Arg Gln Arg Ala Pro Asp Ser Ser Gly His His             |     |     |     |     |
| 166 | 275                                                             | 280 | 285 |     |     |
| 168 | CAC CAT CAT CAC TAATAA GAA TTC GGA TCC TCT AGA GTC GAC AAG CTT  |     |     |     | 912 |
| 169 | His His His Glu Phe Gly Ser Ser Arg Val Asp Lys Leu             |     |     |     |     |
| 170 | 290                                                             | 295 | 300 |     |     |
| 173 | (2) INFORMATION FOR SEQ ID NO: 3:                               |     |     |     |     |
| 175 | (i) SEQUENCE CHARACTERISTICS:                                   |     |     |     |     |
| 176 | (A) LENGTH: 302 amino acids                                     |     |     |     |     |
| 177 | (B) TYPE: amino acid                                            |     |     |     |     |
| 178 | (D) TOPOLOGY: linear                                            |     |     |     |     |
| 180 | (ii) MOLECULE TYPE: protein                                     |     |     |     |     |
| 182 | (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:                        |     |     |     |     |
| 184 | Glu Leu Tyr Lys Ser Gly Asp Pro Gly Glu Asp Pro Arg Leu Glu Pro |     |     |     |     |
| 185 | 1                                                               | 5   | 10  | 15  |     |
| 187 | Trp Lys His Pro Gly Ser Gly Ser Val Asp Pro Asn Leu Glu Pro Trp |     |     |     |     |
| 188 | 20                                                              | 25  | 30  |     |     |
| 190 | Lys His Pro Gly Ser Ser Gly Val Asp Pro Lys Leu Glu Pro Trp Lys |     |     |     |     |
| 191 | 35                                                              | 40  | 45  |     |     |
| 193 | His Pro Gly Ser Ser Gly Val Asp Pro Ser Leu Glu Pro Trp Lys His |     |     |     |     |
| 194 | 50                                                              | 55  | 60  |     |     |
| 196 | Pro Gly Ser Gly Ser Val Asn Pro Ser Leu Glu Pro Trp Lys His Pro |     |     |     |     |
| 197 | 65                                                              | 70  | 75  | 80  |     |
| 199 | Gly Ser Ser Gly Val Asp Pro Arg Leu Glu Pro Trp Glu His Pro Gly |     |     |     |     |
| 200 | 85                                                              | 90  | 95  |     |     |
| 202 | Ser Ser Gly Val Asp Pro Arg Leu Glu Pro Trp Asn His Leu Gly Ser |     |     |     |     |
| 203 | 100                                                             | 105 | 110 |     |     |

RAW SEQUENCE LISTING DATE: 05/07/2002  
 PATENT APPLICATION: US/10/086,208 TIME: 11:26:07

Input Set : N:\Crf3\RULE60\10086208.raw  
 Output Set: N:\CRF3\05072002\J086208.raw

205 Ser Gly Val Asp His Arg Leu Glu Pro Trp Lys His Pro Gly Ser Gly  
 206 115 120 125  
 208 Asp Leu Arg Gln Arg Arg Arg Thr Pro Gln Asp Ser Gly Ser Arg Gln  
 209 130 135 140  
 211 Arg Arg Arg Pro Pro Gln Asp Ser Ser Gly Arg Gln Arg Arg Arg Pro  
 212 145 150 155 160  
 214 Pro Gln Gly Ser Gly Ser Arg Gln Arg Arg Gly Pro Pro Gln Gly Ser  
 215 165 170 175  
 217 Gly Ser Arg Gln Arg Arg Arg Pro Pro Gln Asn Ser Ser Gly Arg Gln  
 218 180 185 190  
 220 Arg Arg Arg Ser Pro Gln Asp Ser Gly Gly Arg Gln Arg Arg Arg Ser  
 221 195 200 205  
 223 Pro Gln Asn Ser Gly Gly Arg Gln Arg Arg Arg Thr Pro Gln Ser Ser  
 224 210 215 220  
 226 Ser Gly Arg Gln Arg Arg Arg Ala His Gln Asn Ser Gly Ser Arg Gln  
 227 225 230 235 240  
 229 Arg Arg Arg Ala His Gln Asp Ser Ser Gly Arg Gln Arg Arg Arg Ala  
 230 245 250 255  
 232 Pro Glu Asp Ser Gly Ser Arg Gln Arg Arg Arg Ala Pro Pro Asp Ser  
 233 260 265 270  
 235 Ser Gly Arg Gln Arg Gln Arg Ala Pro Asp Ser Ser Ser Gly His His  
 236 275 280 285  
 238 His His His Glu Phe Gly Ser Ser Arg Val Asp Lys Leu  
 239 290 295 300

241 (2) INFORMATION FOR SEQ ID NO: 4:

243 (i) SEQUENCE CHARACTERISTICS:  
 244 (A) LENGTH: 9 amino acids  
 245 (B) TYPE: amino acid  
 246 (C) STRANDEDNESS:  
 247 (D) TOPOLOGY: linear

249 (ii) MOLECULE TYPE: peptide

254 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

256 Arg Lys Lys Arg Arg Gln Arg Arg Arg  
 257 1 5

259 (2) INFORMATION FOR SEQ ID NO: 5:

261 (i) SEQUENCE CHARACTERISTICS:  
 262 (A) LENGTH: 6 amino acids  
 263 (B) TYPE: amino acid  
 264 (C) STRANDEDNESS:  
 265 (D) TOPOLOGY: linear

267 (ii) MOLECULE TYPE: peptide

272 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

274 Gly Arg Gly Asp Ser Pro  
 275 1 5

277 (2) INFORMATION FOR SEQ ID NO: 6:

279 (i) SEQUENCE CHARACTERISTICS:  
 280 (A) LENGTH: 7 amino acids  
 281 (B) TYPE: amino acid  
 282 (C) STRANDEDNESS:

## RAW SEQUENCE LISTING

PATENT APPLICATION: US/10/086,208

DATE: 05/07/2002

TIME: 11:26:07

Input Set : N:\Crf3\RULE60\10086208.raw  
Output Set: N:\CRF3\05072002\J086208.raw

283                 (D) TOPOLOGY: linear  
285         (ii) MOLECULE TYPE: peptide  
290         (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:  
292                 Val Asp Pro Arg Leu Glu Pro  
293                 1                         5  
295 (2) INFORMATION FOR SEQ ID NO: 7:  
297         (i) SEQUENCE CHARACTERISTICS:  
298                 (A) LENGTH: 7 amino acids  
299                 (B) TYPE: amino acid  
300                 (C) STRANDEDNESS:  
301                 (D) TOPOLOGY: linear  
303         (ii) MOLECULE TYPE: peptide  
308         (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:  
310                 Val Asp Pro Lys Leu Glu Pro  
311                 1                         5  
313 (2) INFORMATION FOR SEQ ID NO: 8:  
315         (i) SEQUENCE CHARACTERISTICS:  
316                 (A) LENGTH: 7 amino acids  
317                 (B) TYPE: amino acid  
318                 (C) STRANDEDNESS:  
319                 (D) TOPOLOGY: linear  
321         (ii) MOLECULE TYPE: peptide  
326         (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:  
328                 Val Asp Pro Ser Leu Glu Pro  
329                 1                         5  
331 (2) INFORMATION FOR SEQ ID NO: 9:  
333         (i) SEQUENCE CHARACTERISTICS:  
334                 (A) LENGTH: 7 amino acids  
335                 (B) TYPE: amino acid  
336                 (C) STRANDEDNESS:  
337                 (D) TOPOLOGY: linear  
339         (ii) MOLECULE TYPE: peptide  
344         (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:  
346                 Val Asp Pro Asn Leu Glu Pro  
347                 1                         5  
349 (2) INFORMATION FOR SEQ ID NO: 10:  
351         (i) SEQUENCE CHARACTERISTICS:  
352                 (A) LENGTH: 6 amino acids  
353                 (B) TYPE: amino acid  
354                 (C) STRANDEDNESS:  
355                 (D) TOPOLOGY: linear  
357         (ii) MOLECULE TYPE: peptide  
360         (ix) FEATURE:  
361                 (A) NAME/KEY: Binding-site  
362                 (B) LOCATION: 6  
363                 (D) OTHER INFORMATION: /note= "an amide is attached to the  
364 Ser in position 6"  
367         (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:  
369                 Trp Lys His Pro Gly Ser

RAW SEQUENCE LISTING ERROR SUMMARY                   DATE: 05/07/2002  
PATENT APPLICATION: US/10/086,208                   TIME: 11:26:08

Input Set : N:\Crf3\RULE60\10086208.raw  
Output Set: N:\CRF3\05072002\J086208.raw

Please Note:

Use of n and/or Xaa have been detected in the Sequence Listing. Please review the Sequence Listing to ensure that a corresponding explanation is presented in the <220> to <223> fields of each sequence which presents at least one n or Xaa.

Seq#:36; Xaa Pos.1,5,9  
Seq#:37; Xaa Pos.1,3,13  
Seq#:38; Xaa Pos.1,4,5,6,7  
Seq#:39; Xaa Pos.1,4,7,14  
Seq#:58; Xaa Pos.4  
Seq#:86; Xaa Pos.4  
Seq#:87; Xaa Pos.2  
Seq#:88; Xaa Pos.3,4,5,6  
Seq#:89; Xaa Pos.3,6  
Seq#:110; Xaa Pos.3,6  
Seq#:124; Xaa Pos.6

VERIFICATION SUMMARY  
PATENT APPLICATION: US/10/086,208

DATE: 05/07/2002  
TIME: 11:26:08

Input Set : N:\Crf3\RULE60\10086208.raw  
Output Set: N:\CRF3\05072002\J086208.raw

L:27 M:220 C: Keyword misspelled or invalid format, [(A) APPLICATION NUMBER:]  
L:28 M:220 C: Keyword misspelled or invalid format, [(B) FILING DATE:]  
L:875 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:36 after pos.:0  
L:926 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:37 after pos.:0  
L:988 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:38 after pos.:0  
L:1044 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:39 after pos.:0  
L:1400 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:58 after pos.:0  
L:1928 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:86 after pos.:0  
L:1951 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:87 after pos.:0  
L:1992 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:88 after pos.:0  
L:2021 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:89 after pos.:0  
L:2453 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:110 after pos.:0  
L:2720 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:124 after pos.:0